INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v14i2.1103
VOL. 14. NO. 2 FEBRUARI 2020
P-ISSN: 0216-6933 | E-ISSN: 2685-807X

INTEGRASI METODE SAMPLE BOOTSTRAPPING DAN WEIGHTED PRINCIPAL COMPONENT ANALISYS (PCA) UNTUK MENINGKATKAN PERFORMA NAÏVE BAYES PADA CITRA TUNGGAL PAP SMEAR

Yumi Novita Dewi[1]; Harsih Rianto[2]; Dwiza Riana[3]; Juarni Siregar[4]

[1,3,4]Program Studi Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
www.nusamandiri.ac.id
yumi.ymd@nusamandiri.ac.id, dwiza@nusamandiri.ac.id, juarni.jsr@nusamandiri.ac.id

[2]Program Studi Sistem Informasi
Universitas Bina Sarana Informatika
www.bsi.ac.id
harsih.hhr@bsi.ac.id

Abstract—*Research on cervical cancer with the Pap Smear method is useful for finding pre-cancer diagnoses. Associated with previous research that the accuracy of the Naïve Bayes algorithm to the classification of a single Pap smear image still has an unsatisfactory accuracy. Whereas determining the class of single Pap cell smears is very important in determining whether these cells are normal or not. This study aims to determine whether integration using the Sample Bootstrapping (SB) method with the Weighted Principal Component Analysis (W-PCA) algorithm can improve the performance of the Naïve Bayes algorithm for seven different cell types. This model is the best solution used in the classification of datasets that are classified as having large dimensions. So that the integration of the two algorithms can increase the accuracy value to 87.24% for the seven classes and 97.30% for the two classes, and it can be concluded that with this integration model can improve the best accuracy value.*

Keywords*: Pap smear images, Classification, Naïve Bayes, Sample Bootstrapping, Weighted - Principal Component Analysis (PCA).*

Abstrak—Penelitian tentang kanker serviks dengan metode Pap Smear berguna untuk menemukan diagnosa pra-kanker. Terkait dengan penelitian sebelumnya bahwa nilai akurasi algoritma *Naïve Bayes* terhadap klasifikasi citra tunggal Pap smear masih memiliki akurasi yang belum memuaskan. Sedangkan penentuan kelas sel citra tunggal Pap smear sangat penting dalam menentukan apakah sel-sel tersebut normal atau tidak. Penelitian ini bertujuan untuk menentukan apakah dengan cara integrasi menggunakan metode *Sample Bootstrapping* (SB) dengan algoritma *Weighted Principal Component Analysis* (W-PCA) dapat meningkatkan kinerja algoritma Naïve Bayes terhadap tujuh jenis sel yang berbeda. Model ini adalah solusi terbaik yang digunakan dalam klasifikasi dataset yang tergolong berdimensi besar. Sehingga dengan integrasi kedua algoritma tersebut dapat meningkatkan nilai akurasi menjadi 87,24% untuk tujuh kelas dan 97,30% untuk dua kelas, dan dapat disimpulkan bahwa dengan model integrasi ini mampu meningkatkan nilai akurasi terbaik.

Kata kunci: *Pap smear images, Classification, Naïve Bayes, Sample Bootstrapping, Weighted - Principal Component Analysis (PCA).*

PENDAHULUAN

Diperkirakan lebih dari 270.000 kematian akibat kanker serviks setiap tahun. Kanker serviks itu menyerang pada bagian leher serviks yang disebabkan oleh *Human Papilloma Virus* (HPV) yang tidak secara langsung terasa oleh penderitanya, sehingga rata-rata penderita penyakit kanker serviks ini baru akan mengetahui ketika sudah mencapai stadium lanjut. Tentunya hal ini akan menyebabkan kematian bagi

penderitanya jika tindakan medis terlambat (Dewi, Riana, & Mantoro, 2017).

Awal penelitian tentang deteksi dini penyakit kanker serviks dilakukan oleh ilmuan bernama Georgeus Papanicolau pada tahun 1930, dimana Georgeus Papanicolau menemukan mekanisme diagnosa pra-kanker serviks, sehingga deteksi dini penyakit kanker serviks itu dikenal dengan istilah *Pap Smear* (Riana, 2010).

Informasi terhadap data sel kanker serviks tersebut dikenal dengan istilah data herlev oleh peneliti bernama Jantzen, Norup, Dounias, & Bjerregaard, dimana dataset tersebut dapat dilakukan penelitian lebih lanjut tentang perkembangan penyakit kanker serviks (Dewi et al., 2017).

Pada penelitian sebelumnya, hasil segmentasi sel citra *Pap smear* digunakan sebagai input pada sistem untuk mengenali apakah diagnosa sel *Pap smear* yang diinput tersebut sehat atau tidak, dan mengusulkan 20 fitur yang dapat diekstraksi dari hasil segmentasi sel citra tunggal *Pap Smear* (Dewi & Sariasih, 2019). *Jumlah objek klasifikasi sel citra tunggal Pap smear tersebut terdiri atas tujuh kategori kelas. Tiga kelas diantaranya adalah kategori kelas sel normal yang meliputi Normal Superficial, Normal Intermediate, dan Normal Columnar, sedangkan empat yang lainnya adalah kategori kelas sel abnormal yaitu: Mild (Light) Dyplasia, Moderate Dysplasia, Severe Dysplasia dan Carcinoma In Situ* (Riana, 2010).

Dalam perkembangannya, klasifikasi langsung juga ditujukan untuk mengetahui bentuk klasifikasi antara sel normal dan sel tidak normal. Sampai saat ini, lingkupan pengklasifikasian masih terus dilakukan peneliti untuk mengetahui lebih jauh perkembangan penyakit kanker serviks dengan konsep *data mining* (Riana, Widyantoro, & Mengko, 2015).

Data mining adalah proses penggalian pengetahuan dari data yang besar dari basis data atau *repository data base* lainnya. Terkait dengan penelitian sebelumnya, "*Improving Naive Bayes Performance in single Image Pap Smear Using Weighted Principal Component Analysis (WPCA)*", dalam penelitian tersebut telah dilakukannya eksperimen Algoritma Naïve Bayes dapat ditampilkan dalam pengukuran akurasi untuk dataset dari dua kelas karena algoritma Naïve Bayes akan dapat dengan cepat dan akurat menangani dataset dalam dimensi skala kecil sementara uji dataset Naïve Bayes dengan dimensi skala besar mengalami penurunan dalam nilai akurasi. Solusi dari masalah ini adalah mengabungkan antara algoritma Naïve Bayes dan model Principal Component Analysis (PCA) yang digunakan dalam menangani dataset dengan dimensi skala besar sehingga nilai akurasi dapat ditingkatkan menjadi 67,45% dan 87,28% (Dewi et al., 2017).

Untuk meningkatkan akurasi terbaik, perlu adanya intergasi beberapa model algoritma lainnya, sehingga pengklasifikasian dengan dataset sel citra tunggal *Pap Smear* dapat menghasilkan nilai akurasi terbaik. Dengan faktor bahwa dataset sel citra tunggal *Pap Smear* dikategorikan dalam dataset dengan dimensi skala besar, perlu adanya teknik untuk mengurangi jumlah data training untuk diproses dan mengurangi atribut sehingga mampu meningkatkan akurasi dan meminimalkan waktu komputasi.

Penelitian lainnya adalah "Metode *Sample Bootstrapping* Untuk Meningkatkan Performa Algoritma *Naive Bayes* Pada Citra Tunggal *Pap Smear*", dalam penelitian tersebut menggambarkan apakah penggunaan metode *sample bootstrapping* dapat meningkatkan kinerja algoritma naive bayes untuk mengklasifikasikan citra tunggal *pap smear* yang ada pada dataset herlev. Nilai akurasi akan diperiksa untuk dua kelas dan tujuh kelas. Metode yang digunakan terdiri dari beberapa tahapan yaitu *preprocessing*, *knowledge rule*, *evaluation*, dan *performance report*. Hasil penelitian ini menunjukkan bahwa metode *sample bootstrapping* dapat meningkatkan nilai akurasi tujuh kelas menjadi 85,24% dan 93,24% untuk nilai akurasi dengan dua kelas (Dewi & Sariasih, 2019).

Metode *Sample Bootstrapping* digunakan untuk mengurangi jumlah data *training* yang akan diproses. Untuk dapat mengatasi persoalan dengan dataset dengan dimensi besar, maka perlu perlu adanya sampel data (*sampling*) secara acak agar data yang akan diproses menjadi lebih kecil dan untuk mengurangi atribut dalam mengolah data yang besar maka dapat menggunakan metode *Principal Component Analysis* (PCA). Namun PCA memiliki kekurangan dalam kemampuan memilih fitur yang tidak relevan dari dataset. Dengan menggunakan metode *Weighted Principal Component Analysis* dapat mengurangi waktu komputasi sehingga efisien untuk menangani dataset yang memiliki dimensi yang besar (Dewi et al., 2017).

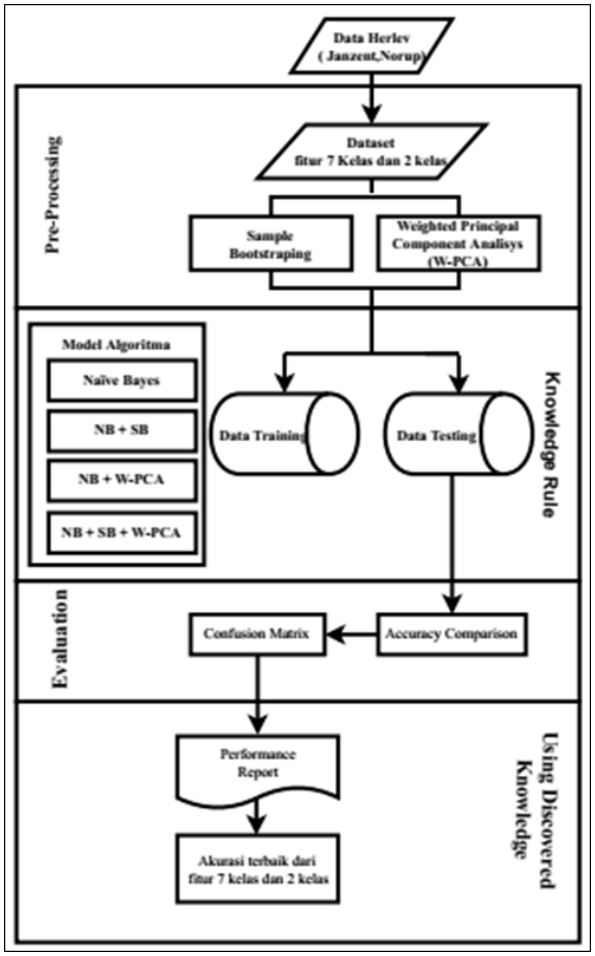
Mengatasi persoalan untuk meningkatkan nilai akurasi terhadap algoritma *naïve bayes* pada dataset sel citra tunggal *Pap Smear*.

BAHAN DAN METODE

Data yang digunakan merupakan dataset sel tunggal *Pap Smear* atau sering juga disebut dengan bank data *Herlev* (Riana, 2010). Dataset ini merupakan data bertipe numerik, dimana dataset sel tunggal *Pap Smear* ini memiliki 20 *fiture attribute*, terbagi menjadi 7 kelas, yaitu: kelas normal dan kelas abnormal sebanyak 917 data*set*.

VOL. 14. NO. 2 FEBRUARI 2020
P-ISSN: 0216-6933 | E-ISSN: 2685-807X

Tiga kelas diantaranya adalah kategori kelas sel normal yang meliputi: *Normal Superficial* (NS), *Normal Intermediate* (NI), dan *Normal Columnar* (NC) (Riana, Plissiti, et al., 2015). Sedangkan empat kelas lainnya adalah kategori kelas abnormal, yaitu: *Mild (Light) Dyplasia* (MLD), *Moderate Dysplasia* (MD), *Severe Dysplasia* (SD), dan *Carcinoma In Situ* (CIS) (Riana, Plissiti, et al., 2015).



Sumber: (Dewi, Rianto, Riana, & Siregar, 2019)

Gambar 1. Rancangan Skema Usulan

Penelitian ini dilakukan dengan mengusulakan model, melakukan eksperimen dengan menguji model yang diusulkan, evaluasi dan validasi model yang diusulkan. Rancangan skema usulan seperti tampak pada Gambar 1. Pada rancangan skema usulan dataset *Harlev* terlebih dahulu dilakukan proses Pre-processing menggunakan algoritma *Sample Bootstraping.* Selanjutnya dataset yang sudah mengalmi proses pre-processing akan dibagi menjadi 10 bagian menggunakan *10-Fold Cross Validation,* data dibagian pertama menjadi data testing dan bagian kedua sampai dengan data bagian kesepuluh menjadi data training. Setelah data terbagi manjadi data training dan data testing dilakukan pengujian model menggunakan algoritma *Naïve Bayes*, *Naïve Bayes* plus *Sample Bootstraping* (NB+SB), *Naïve Bayes* plus *Weighted Principal Component Analysis* (NB+W-PCA) dan *Naïve Bayes* plus *Sample Bootstraping* plus *Weighted Principal Component Analysis* (NB+SB+W-PCA). Hasil dari pengujian model akan dilakukan evalusi menggunakan *Confusion Matrix*, *Accuracy Comparison dan Uji-t atau uji t-test*. Tahapan setelah dilakukan evaluasi adalah menampilkan dalam bentuk laporan performa model.

A. Metode *Sample Bootstraping*

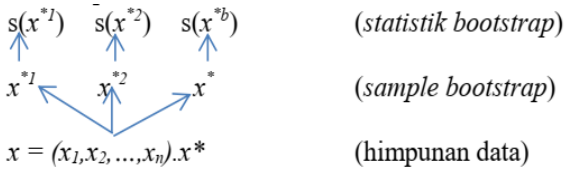
Bootstrapping adalah suatu metode untuk mendeviasikan estimasi yang kuat dari eror standar dan interval kepercayaan untuk mengestimasi proporsi, rata-rata, *median*, *odds ratio*, *koefisien korelasi* atau *koefisien regresi*(Firtiyani & Wahono, 2015). *Bootstrapping* juga dapat digunakan untuk mengembangkan uji hipotesis. *Bootstrapping* sangat berguna sebagai alternatif untuk estimasi parameter ketika peneliti merasa ragu dapat memenuhi asumsi pada data dengan dimensi skala besar ataupun kasus *heteroskedastisitas* yang muncul pada analisis regresi karena ukuran sampel yang dimiliki kecil. Metode *bootstrap* mengetengahkan masalah *resampling* yaitu pengambilan *sample* acak dari *sample* yang sudah ada dengan pengembalian. Secara umum metode *resampling* mengaju pada metode yang menggunakan data amatan secara berulang-ulang dalam suatu analisis simulasi untuk menarik kesimpulan yang dilakukan dengan bantuan *computer* (Zhang et al., 2017).

Misalkan himpunan data asli $x = (x_1,x_2,\ldots,x_n)$.x^* merupakan *sample bootstrap* yang diambil secara acak sebanyak *n* dari data asli dengan pengembalian. Misalkan *s(x)* adalah statistik dari *sample* dari *sample* data asli, maka *s(x)* adalah taksiran nilai data asli, pengambilan *sample bootstrap* dilakukan sebanyak *b* dan nilai taksiran parameter atau statistik diproleh dari nilai statistik dari kumpulan *statistic* $s(x^{*i})$ *dimana i = 1,2, …, b*(Riana, Plissiti, et al., 2015).

Berbeda dengan penaksiran dalam statistika parametrik, metode *sample bootstrap* tidak memerlukan macam-macam asumsi. Satu-satunya yang diperlukan hanyalah bahwa *sample* yang digunakan sudah cukup mewakili (*representative*) populasinya. Disamping itu, metode ini memberikan kemudahan penerapan untuk hampir semua jenis statistic (Pinto Da Costa, Alonso, & Roque, 2011). Banyaknya *sample bootstrap* yang harus diambil idealnya adalah $b \infty$(Zhang et al., 2017). Sedangkan lamanya *computer* memproses akan meningkatkan secara *liner* seiring dengan bertambahnya jumlah *b*. Pengambilan ukuran *sample* acak untuk *bootstrap* yang sudah dianggap

dapat berkisar antara 1.000 sampai 2.000 ulangan(Zhang et al., 2017).

Metode *sample bootstrap* lebih luas penerapannya yaitu dapat digunakan pada *sample* dengan ukuran kurang dari 16 ($n \leq 15$)(Susetyoko & Purwantini, 2010). Metode ini digunakan pada masalah-masalah yang tidak biasa dimana tidak mungkin atau sulit menduga keragaman statistik. Adapun skema proses *sample bootstrap* adalah:

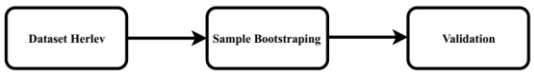


Sumber: (Dewi & Sariasih, 2019)
Gambar 2. Skema Proses *Bootstrap*

Metode *sample bootstrap* dapat juga diterapkan pada hampir semua pendugaan seperti: tidak hanya pada data tunggal *x* tapi juga pada data lebih dari satu atau berpasangan seperti *regresi*, *matrix*, dan *vector*(Zhang et al., 2017). Statistik $t(x)$ yang akan diduga dapat apa saja selama dapat apa saja selama dapat hitung penduga tersebut dari *sample bootstrap* $t(x)$(Zhang et al., 2017). Data tidak harus berasal dari sebarang peluang tertentu dan dapat digunakan pada analisis *regresi*, deret waktu, dan analisis lain yang salah satunya adalah analisis tabel kontigensi. Ukuran keakuratan yang dapat digunakan selain *standart error* adalah *bias*, *mean absolute deviation*, dan selang kepercayaan(Setiawan, Wahono, & Syukur, 2015). Tahapan algoritma *Sample Boostraping* sebagai berikut:

1. Pilih Parameter Sample: *relative* (sample dibuat menjadi sebagian kecil dari jumlah total sample data)
2. Nilai sample ratio diinput antara 0-1.
3. Nilai data bootstrap divalidasi dengan 10-fold cross validation.

Setelah dilakukan sampling, maka data *bootstrap* tersebut divalidasi dengan *10-fold cross validation* sebagaimana ditunjukkan pada Gambar 3:



Sumber: (Dewi et al., 2019)
Gambar 3. Tahapan Pengujian *Naïve Bayes* dengan *Sample Bootstrapping*

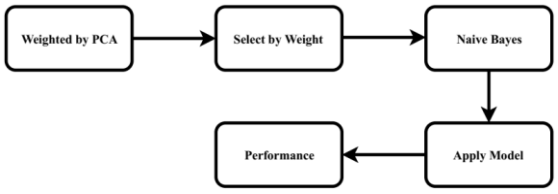
Langkah Selanjutnya adalah melakukan normalisasi kelas data pada dataset dan melakukan pembobotan terhadap atribut kelas dengan *Weighted Principal Component Analysis*.

B. Weighted Principal Component Analysis

Principal Component Analysis adalah salah satu fitur ekstraksi (reduksi) variable yang banyak digunakan (Zhang et al., 2017). Metode Principal Component Analysis sangat berguna digunakan jika data yang ada memiliki jumlah variabel yang besar dan memiliki korelasi antar variabelnya. Perhitungan dari Principal Component Analysis didasarkan pada perhitungan nilai weigen dan vektor eigen yang menyatakan penyebaran data dari suatu dataset (Zhang et al., 2017). Tujuan dari analisa Principal Component Analysis adalah untuk mereduksi variabel yang ada menjadi lebih sedikit tanpa harus kehilangan informasi yang termuat dalam data asli/awal (Khoshgoftaar, Van Hulse, & Napolitano, 2011).

Dengan menggunakan *Principal Component Analysis*, variabel yang tadinya sebanyak *n* variabel akan direduksi menjadi *k* variabel baru (*principal component*) dengan jumlah *k* lebih sedikit dari *n* dan dengan hanya menggunakan *k principal component* akan menghasilkan nilai yang sama dengan menggunakan *n* variable (Susetyoko & Purwantini, 2010). Variabel hasil dari reduksi tersebut dinamakan *principal component* atau bisa juga disebut faktor. Sifat dari variabel baru yang terbentuk dengan analisa *Principal Component Analysis* nantinya selain memiliki jumlah variabel yang berjumlah lebih sedikit tetapi juga menghilangkan korelasi antar variabel yang terbentuk (Zhang et al., 2017).

Langkah selanjutnya adalah melakukan normalisasi terhadap atribut kelas yang mencerminkan relevansi bobot atribut dengan nilai atribut kelas (Zhang et al., 2017). Adapun proses *Weighted Principal Component Analysis* adalah:



Sumber: (Dewi et al., 2019)
Gambar 4. Langkah Pengujian *Naïve Bayes* dengan *Weighted-Principal Component Analysis*

Dalam Pembobotan, nilai akurasi dapat dilihat dari tabel confusion matrix dengan rumus:

$$Accuracy = \frac{Jumlah\ Data\ Benar}{Jumlah\ Data} \times 100\% \text{.....(1)}$$

C. Algoritma Naïve Bayes

Metode klasifikasi yang paling popular dan yang banyak digunakan adalah metode *naïve bayes* (Zhang et al., 2017). Metode *Naive Bayes* merupakan salah satu metode klasifikasi yang dapat memprediksi *probabilitas* keanggotaan dari

VOL. 14. NO. 2 FEBRUARI 2020
P-ISSN: 0216-6933 | E-ISSN: 2685-807X

suatu *class* (Riana, Plissiti, et al., 2015). Nilai dari suatu *class* pada metode *Naive Bayes* bersifat *independen*, dan tidak bergantung pada atribut-atribut lainnya. Klasifikasi ini dilakukan dengan menggunakan rumus sebagai berikut:

$$g(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{.....(2)}$$

Dimana:

$$\mu = \frac{\sum_{i=1}^{n} xi}{n} \quad \text{.....(3)}$$

$$\sigma = \frac{\sqrt{\sum_{i=1}^{n}(x_{i-\mu)^2}}}{n-1} \quad \text{.....(4)}$$

Perhitungan dengan tipe data numerik pada algoritma naïve bayes harus dilakukan dengan perhitungan mean μ dan standart deviation σ menggunakan persamaan (1), (2), dan (3). Semua bagian dilakukan pengujian sampai semua dataset dapat dibagi menjadi dua data yaitu; data training dan data testing (Riana, Plissiti, et al., 2015).

Setiap metode dan algoritma memiliki karakteristiknya masing-masing, begitu juga dengan algoritma naïve bayes (McRoberts, Magnussen, Tomppo, & Chirici, 2011). Klasifikasi naïve bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas (Riana, Plissiti, et al., 2015). Hubungan antara klasifikasi, korelasi hipotesis dan bukti dengan klasifikasi pada naïve bayes adalah label kelas yang menjadi target pemetaan dalam klasifikasi naïve bayes merupakan hipotesisnya, dan fitur-fitur yang menjadi inputan kelas tersebut adalah buktinya (Zhang et al., 2017).

Hal ini memberikan karakteristik tersendiri bagi naïve bayes, adapun karakteristik tersebut adalah:

a) Metode naïve bayes bekerja teguh (robust) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (outliner). Naive bayes juga bisa menangani nilai attribut yang salah dengan megabaikan data training selama proses pembangunan model dan prediksi.
b) Tangguh menghadapi attribut yang tidak relevan.
c) Attibut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi naïve bayes karena asumsi independensi attribut tersebut sudah tidak ada.

HASIL DAN PEMBAHASAN

Algoritma *naïve bayes* dengan metode *sample bootstrapping* merupakan metode yang tepat untuk menaikan nilai *accuracy* terhadap dataset tujuh kelas dan dua kelas sel tunggal *Pap Smear*.

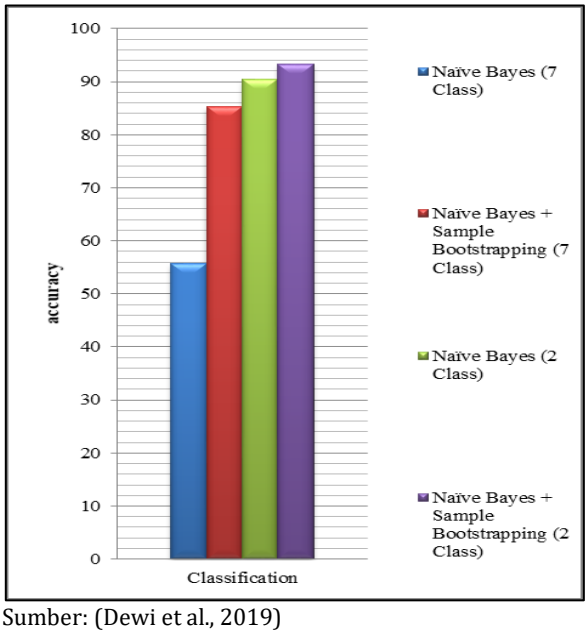
Tabel 1. Perbandingan Algoritma NaiveBayes dengan Sample Boosting

Model Algoritma	Hasil Akurasi	*Kelas*
Naïve Bayes	55,73%	7 Class
Naïve Bayes	90,42%	2 Class
Naïve Bayes + Sample Bootstrapping	85,24%	7 Class
Naïve Bayes + Sample Bootstrapping	93,24%	2 Class

Sumber: (Dewi et al., 2019)

Tabel 1 memperlihatkan perbandingan nilai akurasi pada algoritma *naïve bayes* dan model *naïve bayes* dengan *sample bootstrapping*. Dimana hasil akurasi dari algoritma *naïve bayes* dengan fitur tujuh kelas lebih kecil dari pada dengan model algoritma lainnya.

Berikut ini merupakan grafik dari Tabel perbandingan algoritma *Naïve Bayes* dengan *Sample Bootstrapping* yang menunjukkan bahwa *chart* tertinggi ditunjukkan oleh hasil komparasi algoritma *naïve bayes* dengan model *sample bootstrapping*.



Sumber: (Dewi et al., 2019)

Gambar 5. Grafik perbandingan NB dengan SB

Selanjutnya akan dilakukan uji-*t* guna melihat apakah peningkatan yang didapatkan memberikan pengaruh yang cukup signifikan terhadap data.

Tabel 2. Uji-t 7 Kelas dan 2 Kelas Terhadap NB dan SB

	0,872	0,872
0,872	**-**	**1,000**
0,872	-	-

Sumber: (Dewi et al., 2019)

Hasil uji-*t* dari algoritma *Naive Bayes* model *Sampe Bootstrappin*g terhadap dataset tujuh kelas dan dua kelas adalah 1,000. Angka ini lebih besar dari nilai α yaitu 0,05. Artinya, tidak ada perbedaan yang signifikan terhadap keduanya.

Metode *sample bootstrapping* digunakan untuk mengurangi jumlah data *training* yang akan diproses. Selanjutnya akan dilakukan proses pengurangan *attribute* dengan mengusulkan penggunakan metode *Weighted Principal Component Analisis* (PCA) sebagai model untuk meningkatkan akurasi yang optimal pada algoritma *naive bayes*, khususnya terhadap dataset tujuh *class* citra tunggal *pap smear*. Berikut ini merupakan tabel perbandingan dari algoritma algoritma *naïve bayes* dengan model *Weighted-Principal Component Analisis* (*PCA*).

Tabel 3.Perbandingan Naive Bayes dengan W-PCA

Model Algoritma	Hasil Akurasi	*Kelas*
Naïve Bayes	55,73%	7 Class
Naïve Bayes	**90,42%**	2 Class
Naïve Bayes + Weighted-Principal Component Analisis (PCA)	**87,24%**	7 Class
Naïve Bayes + Weighted-Principal Component Analisis (PCA)	67,45%	2 Class

Sumber: (Dewi et al., 2019)

Tabel 4. Memperlihatkan perbandingan nilai akurasi pada algoritma *naïve bayes* dan model *naïve bayes* dengan model *Weighted-Principal Component Analisis* (*PCA*). Dimana hasil akurasi dari algoritma *naïve bayes* dengan fitur tujuh kelas lebih kecil dari pada dengan model algoritma lainnya. Akan tetapi pada kombinasi algoritma *naïve bayes* dengan model *Weighted-Principal Component Analisis* (*PCA*) pada fitur dua kelas mengalami penurunan akurasi menjadi 67,45%.

Berikut ini merupakan Gambar 3 grafik dari tabel perbandingan algoritma *Naïve Bayes* dengan model *Weighted-Principal Component Analisis* (*PCA*) yang menunjukkan bahwa *chart* tertinggi ditunjukkan oleh hasil akurasi algoritma *naïve bayes* dengan fitur dua kelas.

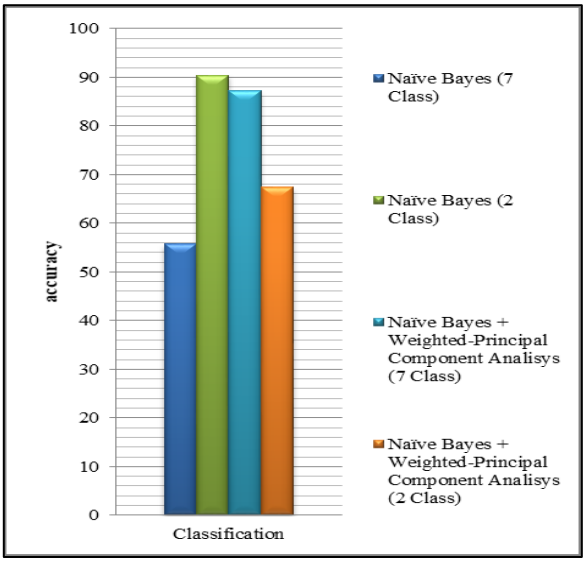
Selanjutnya akan dilakukan uji-*t* guna melihat apakah peningkatan yang didapatkan memberikan pengaruh yang cukup signifikan terhadap data.

Tabel 4. Uji-t 7 Kelas dan 2 Kelas Teradap NB dan W-PCA

	0,939	0,959
0,939	**0,040**	-
0,959	-	-

Sumber: (Dewi et al., 2019)

Hasil uji-*t* dari algoritma *Naive Bayes* dengan model *Weighted-Principal Component Analisis* (*PCA*) terhadap dataset tujuh kelas dan dua kelas adalah 0,040. Angka ini lebih kecil dari nilai α yaitu 0,05. Artinya, kedua model tersebut memiliki perbedaan yang signifikan terhadap kedua kelas sel tunggal *Pap Smear*.



Sumber: (Dewi et al., 2019)

Gambar 6. Grafik Perbandingan NB dengan W-PCA

Terlihat pada Gambar 6 merupakan hasil perbandingan komparasi model algoritma *Naïve Bayes*, *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) dimana *Naïve Bayes* dengan model ini dapat digunakan untuk menaikan nilai akurasi terhadap dataset tujuh kelas dan dua kelas pada dataset sel tunggal *Pap Smear*.

Pada Tabel 6 dapat dilihat perbedaan nilai akurasi dari hasil pengukuran keseluruhan model algoritma. Secara keseluruhan algoritma *naïve bayes* dapat diunggulkan dalam hal pengukuran akurasi untuk dataset sel tunggal *Pap Smear* dengan dua kelas, karena algoritma *naive bayes* akan mampu dengan cepat dan akurat untuk menangani dataset dengan dimensi kecil. Sedangkan uji dataset *naive bayes* dengan dataset berdimensi besar mengalami menurunan nilai akurasi. Sebaliknya jika dilihat dari dataset sel tunggal *Pap Smear* dengan tujuh kelas mengalami penurunan nilai akurasi. Dan komparasi keseluruhan model algoritma *Naïve Bayes*, *Sample Bootstrapping*, dan *Weighted-Principal Component Analisis* (*PCA*)-lah yang menunjukkan nilai akurasi terbaik.

Tabel 5. Hasil Perbandingan Komparasi Algoritma NB, SB, dan W-PCA

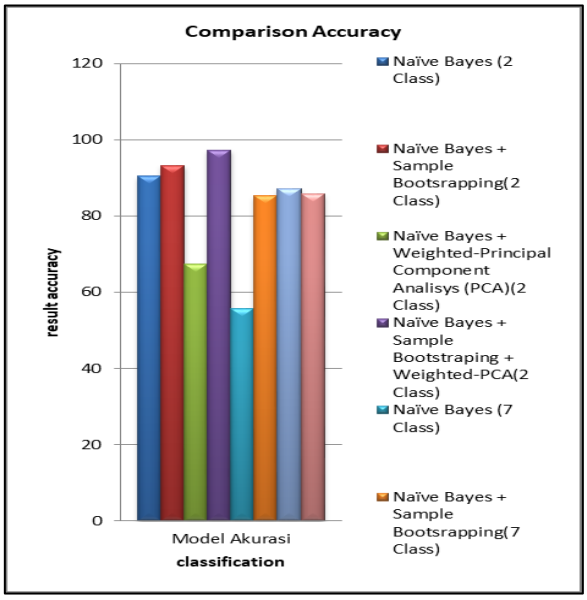
Kelas	Model Akurasi	Hasil
2 *Class*	*Naïve Bayes*	90,42%
	Naïve Bayes + Sample Bootsrapping	93,24%

VOL. 14. NO. 2 FEBRUARI 2020
P-ISSN: 0216-6933 | E-ISSN: 2685-807X

	Naïve Bayes + Weighted-Principal Component Analisis (PCA)	67,45%
	Naïve Bayes + Sample Bootstraping + Weighted-PCA	**97,30%**
7 *Class*	*Naïve Bayes*	55,73%
	Naïve Bayes + Sample Bootsrapping	85,24%
	Naïve Bayes + Weighted-Principal Component Analisis (PCA)	**87,24%**
	Naïve Bayes + Sample Bootstraping + Weighted-PCA	85,70%

Sumber: (Dewi et al., 2019)

Berikut ini merupakan grafik dari tabel perbandingan komparasi model algoritma *Naïve Bayes*, *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) pada dataset dua kelas dan tujuh kelas yang menunjukkan bahwa *chart* tertinggi ditunjukkan oleh hasil akurasi algoritma *naïve bayes* dengan fitur dua kelas.



Sumber: (Dewi et al., 2019)

Gambar 7. Grafik perbandingan algoritma NB, SB, dan W-PCA

Selanjutnya, dilakukan uji-*t* untuk menguji bahwasannya perbedaan antara model *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* tidak terjadi secara kebetulan. Hasil pengujian berbeda dengan menggunakan uji-*t* yang dilakukan dengan aplikasi *rapidminer*. Berikut merupakan tabel hasil uji-*t* algoritma *Naive Bayes* terhadap model *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*).

Hasil uji-*t* dari algoritma *Naive Bayes* dengan model *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) adalah 0,023. Angka ini lebih kecil dari nilai α yaitu 0,05. Artinya, kedua model algoritma *Naive Bayes* dengan model *Sampel Bootstrappin*g dan *Weighted - Principal Component Analisis* (*PCA*) tersebut memiliki perbedaan yang signifikan terhadap kedua kelas sel tunggal *Pap Smear*.

Tabel 6. Uji-t NB Terhadap SB dan W-PCA

	0,857	0,872
0,857	**0,023**	-
0,872	-	-

Sumber: (Dewi et al., 2019)

Solusi dari persoalan kasus ini adalah dengan mengkomparasikan algoritma *Naive Bayes* dengan model *Sampe Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) yang digunakan dalam menangani dataset dengan dimensi skala besar. Sehingga nilai akurasi dapat meningkat menjadi 86,59%. Maka dapat disimpulkan bahwa algoritma *Naive Bayes* dengan model *Sampe Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) mampu meningkatkan akurasi pada dataset sel tunggal *Pap Smear* dengan *rule* tujuh kelas menjadi lebih baik.

KESIMPULAN

Secara keseluruhan algoritma *naïve bayes* dapat diunggulkan dalam hal pengukuran akurasi untuk dataset dua kelas, karena algoritma *naive bayes* akan mampu dengan cepat dan akurat untuk menangani dataset dalam dimensi skala kecil. Sedangkan uji dataset *naive bayes* dengan dimensi skala besar mengalami menurunan nilai akurasi. Solusi dari persoalan kasus ini adalah dengan mengkomparasikan algoritma *naive bayes* dengan model *sample bootstrappin*g dan *weighted - Principal Component Analisis* (*PCA*) yang digunakan dalam menangani dataset dengan dimensi skala besar. Sehingga nilai akurasi dapat meningkat menjadi 86,59% dan 97,30%.

Hasil pengujian dengan uji-*t* dilakukan dengan implementasi *rapidminer* terhadap algoritma *Naive Bayes* dengan model *Sample Bootstrappin*g dan *Weighted-Principal Component Analisis* (*PCA*) dengan prolehan angka sebesar 0,023. Angka ini lebih kecil dari nilai α yaitu 0,05. Artinya, kedua model algoritma *Naive Bayes* dengan model *Sampel Bootstrappin*g dan *Weighted - Principal Component Analisis* (*PCA*) tersebut memiliki perbedaan yang signifikan terhadap kedua kelas sel tunggal *Pap Smear*.

REFERENSI

Dewi, Y. N., Riana, D., & Mantoro, T. (2017). Improving Naïve Bayes performance in single image pap smear using weighted principal component analysis (WPCA). *2017 International Conference on Computing, Engineering, and Design (ICCED)*, *2018-March*, 1–5. https://doi.org/10.1109/CED.2017.8308130

Dewi, Y. N., Rianto, H., Riana, D., & Siregar, J. (2019). *INTEGRASI METODE SAMPLE BOOTSTRAPPING DAN WEIGHTED PRINCIPAL COMPONENT ANALISYS (PCA) UNTUK MENINGKATKAN PERFORMA NAÏVE BAYES PADA CITRA TUNGGAL PAP SMEAR*.

Dewi, Y. N., & Sariasih, F. A. (2019). Metode Sample Bootstrapping Untuk Meningkatkan Performa. *Jurnal Teknik Informatika*, *12*(1), 1–10.

Firtiyani, fitriyani, & Wahono, R. S. (2015). Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes. *Journal of Software Engineering*, *1*(2), 101–108.

Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *41*(3), 552–568. https://doi.org/10.1109/TSMCA.2010.2084081

McRoberts, R. E., Magnussen, S., Tomppo, E. O., & Chirici, G. (2011). Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. *Remote Sensing of Environment*, *115*(12), 3165–3174. https://doi.org/10.1016/j.rse.2011.07.002

Pinto Da Costa, J. F., Alonso, H., & Roque, L. (2011). A Weighted Principal Component Analysis and Its Application to Gene Expression Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(1), 246–252. https://doi.org/10.1109/TCBB.2009.61

Riana, D. (2010). *Hierarchical Decision Approach Berdasarkan Importance Performance Analysis Untuk Klasifikas Citra Tunggal Pap Smear Menggunakan Fitur Kuantitatif dan Kualitatif*. Universitas Indonesia.

Riana, D., Plissiti, M. E., Nikou, C., Widyantoro, D. H., Mengko, T. L. R., & Kalsoem, O. (2015). Inflammatory cell extraction and nuclei detection in pap smear images. *International Journal of E-Health and Medical Communications*, *6*(2), 27–43. https://doi.org/10.4018/IJEHMC.2015040103

Riana, D., Widyantoro, D. H., & Mengko, T. L. (2015). Extraction and Classification Texture of Inflammatory Cells and Nuclei in Normal Pap smear Images. *2015 4th International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering, ICICI-BME*, 65–69. https://doi.org/10.1109/ICICI-BME.2015.7401336

Setiawan, T. A., Wahono, R. S., & Syukur, A. (2015). Integrasi Metode Sample Bootstrapping dan Weighted Principal Component Analysis untuk Meningkatkan Performa K Nearest Neighbor pada Dataset Besar. *Journal of Intelligent Systems*, *1*(2), 76–81.

Susetyoko, R., & Purwantini, E. (2010). *Teknik Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil*. *2010*, 978–979.

Zhang, L., Lu, L., Nogues, I., Summers, R. M., Liu, S., & Yao, J. (2017). DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE Journal of Biomedical and Health Informatics*, *21*(6), 1633–1643. https://doi.org/10.1109/JBHI.2017.2705583

